Integrating psychometric and computational approaches to individual differences in multimodal reasoning

in Proceedings of the 21st Annual Cognitive Science Society Conference, Vancouver: LEA (1999). pp.405-410.

Integrating psychometric and computational approaches to individual differences in multimodal reasoning

Padraic Monaghan (Padraic.Monaghan@ed.ac.uk)
Keith Stenning (K.Stenning@ed.ac.uk)
Jon Oberlander (J.Oberlander@ed.ac.uk)
Human Communication Research Centre, Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK

Cecilia Sönströd (cecilia@www.phil.gu.se)
Department of Philosophy, Gothenburg University, Gothenberg, Sweden

Abstract

Psychometric measures of ability are unsuited to computational descriptions of tasks, primarily because they cannot take process into account. Studies of aptitude–treatment interactions have often failed to replicate from task to task precisely because of this difficulty. The current study aligns psychometric measures with process accounts in the domain of multimodal reasoning. Learning from multimodal logic courses transfers to other reasoning tasks, and this transfer has been found to relate to differences in strategic use of graphical representations in proof construction. The current study is a replication and an extension of these findings. Different goal types are distinguished in terms of: their modality; whether they involve proofs of consequence or non-consequence; and whether they can be solved by constructing single or multiple cases. We report on the interaction of a range of psychometric measures, and the ways in which they relate to the development and deployment of strategies. In particular, students who develop coping strategies to overcome difficulties with certain problems find that these strategies arise at the expense of appropriate use of a variety of strategies. Our approach, which characterises goals in terms of their logical as well as phenomenal properties, supports a computational perspective on psychometric measures in reasoning tasks.

Introduction

The process of learning to construct formal proofs combining diagrammatic and sentential representations provides a unique microcosm for investigating aptitude–treatment interactions (ATIs) based on different representational behaviours (Stenning, Cox & Oberlander, 1995). The formality of the representations and processes offers the possibility of producing computational models of the mental processes involved. Psychometric approaches which posit no accounts of the mental processes which underly their measurements not only block connections to cognitive accounts of mental process, but frequently lead to failure to replicate ATIs (Cronbach & Snow, 1977). Transfer of scientific theory from situation to situation is dependent on accounts of underlying structure and process—just like transfer of students' learning.

This paper seeks to replicate and extend earlier work on ATIs in learning logic from Hyperproof (HP), a multimodal proof environment due to Barwise and Etchemendy (1994). The HP interface is shown in Figure 1. The top of the main window displays a graphical situation. Below the situation are sentential statements that refer to the graphical situation. The other windows indicate goals for which the student has to construct proofs—in this example, there are four different goals, one sentential and three graphical. Graphical situations can contain abstraction in several ways: size (small, medium, large) and shape (tetrahedron, cube, dodecahedron) can be left unspecified via cylinders and bags, respectively; and position of blocks in the situation can be left unspecified by blocks appearing off the board.

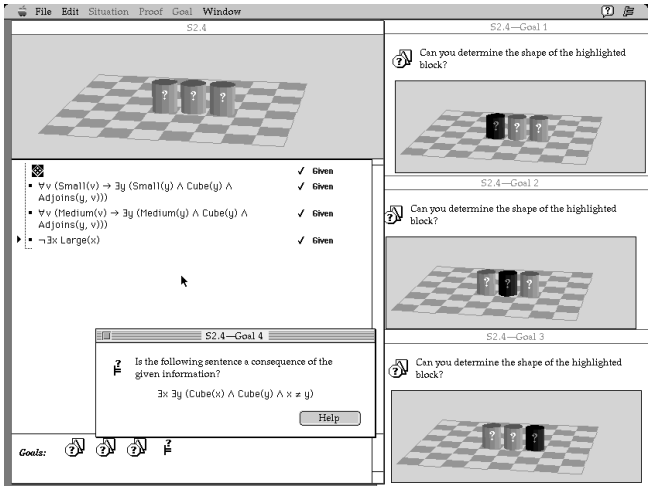

Figure 1: The Hyperproof interface—Problem 4.

An earlier study (Stenning, Cox & Oberlander, 1995) revealed that scores on a subscale of the Graduate Record Exam (GRE) analytical reasoning test predicted outcomes of learning supported by HP with or without its diagrammatic component. Students were classified into high and low scoring on the constraint satisfaction problems of this test. For students taught a 10 week course with HP, those that scored high on the GRE pre-test showed pre- to post-test improvement on a 'blocks-world' (BW) test (see Methods section for a description of this test), whereas those that scored low on the GRE pre-test actually showed decrements on the BW test. These results were reversed when the teaching was a 10 week conventional logic course taught using only the sentential component of HP. Subsequent analysis of the logs of the proofs these students produced in their exams showed that the two groups displayed contrasting proof structures on some problems. These problems were characterised, as predicted, on computational grounds, by the use of a high degree of graphical abstraction (Oberlander et al., 1996).

Monaghan & Stenning (1998) replicated this ATI in the domain of syllogistic reasoning. Subjects who scored high on the GRE constraint satisfaction problems learned a graphical method for solving syllogisms with fewer errors and greater

ease than their lower scoring counterparts. Teaching with a method that was based on sentential natural deduction reversed the effect. The study also showed that other tests of spatial ability correlated with difficulties at different stages of syllogistic reasoning. Thus, in these two domains of study, representation and strategy have been shown to inter-relate: the interplay of style and modality is an important factor in learning to solve reasoning tasks.

As well as aiming to replicate these previous studies with a different student population, we had two further goals. The first was to examine the relationship between several established psychometric measures that deal with 'spatial' and 'verbal' processing, and a multimodal reasoning environment. The second was to deepen our understanding of the dimensions of proof strategy which distinguish students with different reasoning styles.

The earlier study revealed systematic differences in the use of graphical abstraction on one indeterminate exam problem. Here we seek to explore the effects of goal-types in HP differing in (i) modality (sentential versus graphical); (ii) consequentiality (versus non-consequentiality); and (iii) multiplicity of required cases (versus the possibility of a single case being sufficient). For a given goal, there is a range of proof methods available. This classification of goal types and proof methods enables a controlled examination of strategic flexibility in choice of proof methods. The relative contributions of representational and strategic differences is an important issue in the psychometrics of reasoning (see for instance Roberts, 1998), as is the issue of flexibility of approach for psychometrics more generally (Guilford, 1980).

Method

84 students registered on a philosophy degree at the University of Gothenberg participated in the experiment. They followed the HP course material (Barwise & Etchemendy, 1994) as part of a course on introductory logic. The HP coursework was done in parallel with the students learning from more traditional sources: in particular, they learned a traditional natural deduction method of proof (Bennet, Haglund, Westerståhl & Sönströd, 1997), which was based on Mates' (1965) natural deduction method.

At the end of the course, students were set six problems in Hyperproof, which they were free to solve in their own time. These problems were designed to be 'indeterminate', containing a high degree of abstraction in the graphics. One exam question (no. 4) was the same as one used in the original HP study. Proofs were computer-logged, providing detailed data on temporal and ordinal aspects of proof construction.

The students sat a range of pre-tests and one post-test voluntarily. The number of students participating in each stage varied (see Results section).

Pre- and post-course tests

Pre- and post-course tests were used in order to replicate and extend the original HP studies. The same tests as were used in the first HP study were administered, but in addition we use other psychometric measures that are relevant to spatial and verbal information processing.

The GRE test was the same as that used in the first HP study. The test has two types of problem: 'analytical' items (GREA) are those where the construction of a diagram is useful – these are the constraint satisfaction problems where a model can be constructed from the information; and verbal reasoning items (GREV) which require argument analysis, and assessment of the similarity of arguments. For these verbal items, several models may be consistent with the given information. For more details on this test see Cox et al. (1996).

As a measure of transfer of reasoning skills, the Blocks' World test was administered. This is a paper and pencil test which requires students to reason about situations similar to those presented in HP but with natural language descriptions of the conditions on those situations. Different versions of this test were given both before and after the course.

To measure 'spatial ability', the paper folding test (PFT) (French, Ekstrom & Price, 1963) was used. This requires the participant to decide on the array of holes resulting from a piece of paper being folded in various ways, having a hole punched in it, and then unfolded again. This can be interpreted as a measure of strategic flexibility in using spatial information, rather than reflecting spatial ability *per se* (Kyllonen & Lohman, 1983).

Students also took the embedded figures test (EFT). This requires the student to locate geometrical figures in a rectangle of crossing lines. Students who perform well on this task are classified as field-independent, those who perform less well are field-dependent (Witkin et al., 1971). Field-independent students are more likely to process information independently from the context, whereas field-dependent students are more likely to take the context into account.

All tests were set in English, and students were classified according to a median-split on each of the pre-tests.

Results and Discussion

61 students did the EFT and the PFT, 57 students did the pre-course BW test, and 59 students did the GRE test; 72 students completed the HP exam problems; and 27 students did the BW post-course test. Some of the HP records were lost due to bugs in the logging program, so n-values vary throughout the analyses. Full data for all questions and pre-tests exists for 39 students.

Correlations between the pre-test measures (shown in Table 1[1]) were in accord with the literature (Jonassen & Grabowski, 1993): the two subscales of the GRE are correlated; GREA correlates with EFT and PFT scores; and EFT and PFT show a significant, but slight, correlation.

These pre-test measure correlations indicate that the group is representative of a general student population. Analyses of

[1]Numbers in parentheses are the degrees of freedom. * indicates two-tailed significance $p<0.05$.

Table 1: Correlations between pre-tests.

	GREV	EFT	PFT
GREA	0.30* (57)	0.34* (55)	0.27* (55)
GREV		0.22 (55)	0.22 (55)
EFT			0.26* (59)

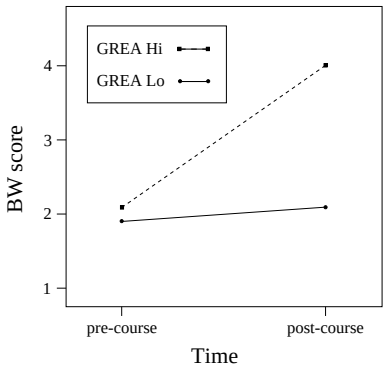

Figure 2: Change in BW score by GREA group.

the differences in generalising *from* and differences in reasoning *within* HP are now reported.

Transfer from HP

To relate the current study to existing results from the original HP experiment, the 'transfer' of reasoning skills from the course to the BW test was measured. Students were divided by a median split into those that scored high and those that scored low on the GREA subscale. Figure 2 illustrates the relationship between pre- and post-test score and GREA group.

As with the original study, those scoring high on the GREA subscale benefitted more from following the HP course ($t(23) = 1.80$, one-tailed $p<0.05$).

Strategies within HP

The original HP studies reported two different responses to graphical abstraction in HP, which related to the GREA groups. GREA high scoring students use more abstraction in their proofs, whereas GREA low students utilise highly concrete situations as they solve the problems. These different strategies were found only on HP problems that contained graphical abstraction (Figure 1 shows an indeterminate problem (no. 4)). As the exam in the current study was deliberately designed to consist only of this sort of problem, we expected a similar dichotomy of strategies to emerge.

The strategies observed in the original HP study can be distinguished by the proportion of fully concrete as opposed to abstract graphical situations that the student creates in a proof (Monaghan, 1998). GREA high students' proofs have more abstract situations, GREA low students' proofs have more concrete situations. However, this distinction does not predict different proofs in the current study. The GREA highs use fully concrete situations 40% of the time, GREA lows construct concrete situations 37% of the time. However, the EFT scores did seem to reflect this stylistic difference better: EFT high scoring students used concrete situations 32% of the time, compared to EFT low scoring students' 51% use, which is significantly more ($t(51) = 2.29$, $p<0.05$).

It may be that the EFT is a better indicator of this strategic difference for this population – answering the GRE test in a second language may mean that scores also reflect language competence. It might be expected that the EFT and the GREA correlate, and in the current study this is the case. However, in the syllogistic study reported earlier (Monaghan & Stenning, 1998) no correlation was found between these measures for native English speakers ($r(20) = 0.03$, $p>0.8$), and in the current study transfer of ability to the BW test was not found to improve more for one EFT group over the other ($t(23) = 0.11$, $p>0.9$).

Differences in the observed strategies used in HP may also be due to differences in teaching in the current study, which combined sentential and multimodal teaching. In the original HP study, when sentential materials alone were used to teach logic, the GREA low group transferred reasoning skills better to the BW test.

This complicated mixture of success and failure to replicate at the level of psychometric tests has bedevilled the ATI literature (see Cronbach & Snow, 1977, for a review). Spatial ability, for example, has been variously defined as ability for the "encoding, transformation, and recognition of spatial information" (Salthouse et al., 1990). The measure suffers from even less precision when strategic variation is considered. Kyllonen & Lohman (1983) show that such tests as the PFT are solved by some students with strategies that seem to invoke representations that are not spatial. This has led Roberts (1998) to consider such measures from the perspective of strategic variation and change, arguing that it is these patterns of strategic variation and change which characterise the psychometric measures.

Our response is to look for more principled ways of analysing proof styles which have clearer relations to computational processes. Ultimately, theoretically motivated process accounts of proof styles and an account of how they generalise across tasks should replace unprincipled test scores. For the present, the psychometric scores provide some empirical reassurance of general applicability. We had expanded the indeterminate problems in the exam used in the present study precisely so that we could explore styles more systematically. HP offers an environment where the student's preference for expressing information *within* a particular modality can be assessed; where the transfer of information *between* modalities can be plotted; and where variation of proof method with problem type can be explored.

In order to explore the issue of strategic flexibility within HP we need two classifications: of goal-types and of proof methods. HP problems can pose a number of goals that have to be solved with reference to one situation. These goals vary in terms of *modality*: the goal can be about a graphical state of affairs, or about a sentential statement (G- or S-goal). Within the graphical modality, goals can vary as to whether the requirement is to prove that a particular situation is a consequence of the given information (graphical-consequence: GC-goal), or to prove that the situation is not a consequence of the given information (graphical-non-consequence: GN-goal). Finally, the GC-goals can be distinguished between those that require splitting into *multiple* situations in order to reach the conclusion (GCM-goal), and those that can be achieved by applying sentential information to a *single* graphical situation (GCS-goal). In HP, the former are usually those that require the position of blocks to be decided (though problem 4 illustrated in Figure 1 is an exception). The latter require the shape or size of the shape to be determined. In problem 4 (Figure 1), goal 4 is an S-goal, goals 1 and 3 are GN-goals, and goal 2 is a GCM goal.

Having chosen a classification of goals, we also need a classification of proof-methods so that we can explore how problem-type and learning style combine to determine proof method. HP is designed to teach 'proof-by-cases'. All goal types can be achieved by constructing a number of graphical situations (cases) and using sentential or graphical rules to prove or exclude the goal for each case. Using fewer situations reflects the use of more abstraction in the way the situations are characterised. This suggests the use of number of cases as a general structural index of proof method. We first seek some empirical support for this measure.

When the pattern of situations that the students produced in their HP proofs was examined more closely, one striking feature emerged: some students used very few graphical situations in their proofs. For one problem in particular (in fact no. 4, the question from the original study), all steps in these students' proofs were rules that operated on sentential information in the sentential window of HP. These proofs maintained the maximum level of abstraction, and so can be seen as being the extreme case of the graphical proof that invokes abstract graphical situations. Thus, three types of proof are distinguishable in the current study, and they can be located, not in terms of abstract or concrete graphical situation construction, but in terms of the mean level of concreteness of the proof. For question 4, where instances of the 'sentential' strategy occur, mean concreteness and EFT score are significantly related ($r(49) = 0.35$, $p<0.02$).

Students were classified as using one of these three proof-types on problem 4, and pre-test scores were compared for these three groups. The results are shown in Table 2.

On semantic grounds, the three strategies constitute points on a continuum—sentential proofs are merely more abstract proofs which happen to be in a different modality. This continuum is reflected in EFT and PFT scores. Those students producing more abstract proofs independent of modality, have high EFT and PFT scores, and they better transfer learned reasoning skills to the post-course BW test. The differences in EFT score between the two extremes of abstraction/concreteness are significantly different ($t(44) = 2.65$, $p<0.02$). The group including both sentential and abstract graphical strategies scored higher on the EFT than the concrete graphical strategy group ($t(49) = 2.64$, $p<0.02$). However, the EFT score for the graphical abstract group alone did not differ significantly from either other strategy group. The ordering of EFT scores across the three proof methods helps justify the use of the number of situations generated in a proof as an index of proof methods across problems.

Table 2: Pre-test scores for students classified by three strategies for Problem 4.

Pre-test	Strategy		
	sentential	abstract graphical	concrete graphical
GREA	5.00	4.25	4.67
GREV	2.00	1.75	2.97
EFT	19.10	16.40	12.56
PFT	13.50	12.80	12.36
BW improvement	1.50	1.00	0.76

The modality independence constituted by sentential proof is consonant with the EFT being a measure of field-dependence–field-independence. When information is presented graphically, students who are more field independent (scoring high on the EFT) are more able to represent that information abstracted from the context it is presented in (Witkin et al., 1971).

This classification of proof methods applied only to the data of problem 4 does not reveal any relation to GRE scores. However, relations reappear when the two classifications of goal-type and proof method are applied to all the data.

Proof methods analysed by goal-types

For each goal type, the number of situations used to achieve the goal was analysed. Two-way ANOVAs were carried out, with number of situations used for each type of goal as a repeated measure and median splits on the pre-tests (EFT, PFT, GREA, GREV) as between-subjects variables. Using these analyses, strategic approach to the different types of goal can be assessed and related to the pre-tests. Three different ANOVAs were carried out: G-goals compared to S-goals; GC-goals compared to GN-goals; and GCM-goals compared to GCS-goals. Only results reaching significance are reported below. Between-subjects effects indicate whether the pre-test alone distinguishes overall differences in the number of situations used to achieve the goals; within-subjects effects measure the interaction between goal type and the pre-tests.

Sentential & Graphical goals

Between subjects effects that proved significant are shown in the first part of Table 3.

This analysis shows that the number of situations used to solve the goals is sensitive to several of the pre-test measures. Those that score higher on the PFT use fewer situations to solve the goals. Those that score higher on both the EFT and the PFT use fewer situations than those that score lower

Table 3: Comparing S and G goals: between-subjects effects and within-subjects effects.

Pre-test(s)	Between-subjects effects
PFT	$F(1, 25) = 8.31$, $p<0.01$
EFT by PFT	$F(1, 25) = 13.01$, $p<0.005$
EFT by GRE-A	$F(1, 25) = 7.05$, $p<0.02$
PFT by GRE-V	$F(1, 25) = 7.49$, $p<0.02$
GRE-A by GRE-V	$F(1, 25) = 4.74$, $p<0.05$
EFT by PFT by GRE-V	$F(1, 25) = 6.40$, $p<0.02$
EFT by GRE-A by GRE-V	$F(1, 25) = 7.61$, $p<0.02$

Pre-test(s)	Within-subjects effects
goal type	$F(1, 25) = 176.99$, $p<0.001$
EFT by goal-type	$F(1, 25) = 4.24$, $p<0.05$
PFT by goal-type	$F(1, 25) = 4.99$, $p<0.05$
EFT by PFT by goal-type	$F(1, 25) = 7.04$, $p<0.02$
PFT by GRE-V by goal-type	$F(1, 25) = 4.74$, $p<0.05$

Table 4: Mean number of situations for goal type by PFT and GREV.

S-goal	PFT Lo	PFT Hi
GREV Lo	10.08	4.89
GREV Hi	7.25	7.33
G-goal	PFT Lo	PFT Hi
GREV Lo	37.63	20.88
GREV Hi	25.25	29.00

on one or both of these measures. However, there is not a simple association between high test scores and efficiency in situation use, as measuring the overall proof length alone is unrelated to any of the psychometric scores.

The second part of Table 3 displays within-subjects effects and these results indicate that students that score differently on the pre-test differ in their solutions for each of the goal types. Both EFT Hi students and PFT Hi students use fewer situations for each type of goal. Here high ability on these measures relates to using fewer situations, but the interaction of PFT by GREV on goal-type points to a more stylistic variation, and one that is glossed over when only a single psychometric is used to distinguish response. Table 4 shows the mean number of situations used for each goal type distinguished by PFT and GREV. There are similar interactions for both types of goal: PFT Hi students use fewer situations for each goal type, but this is modulated by GREV group. If the student is in the GREV high group, then it doesn't matter which PFT group they are in: they use the same number of situations for each goal type. If the student scores low on the GREV, then being PFT Lo means a large number of situations are used for each goal, and being PFT Hi means that the fewest situations are used for each goal.

Students who score high on the GREV scale are those who are good at solving problems that do not require breaking into cases. Scoring high on this scale means that flexibility in using graphical representations (measured by the PFT scale) is irrelevant. Scoring low on the GREV means that students who are good at using graphical representations to support reasoning (PFT Hi) utilise the graphical abstraction facilities of HP to their full potential. Those who are low on both scales rely more on concretising the problem's information.

Graphical goals: consequence & non-consequence

For between-subjects effects, the results were identical to the S- and G-goal analysis. EFT Hi and PFT Hi students used fewer situations for both types of problem. However, the only within-subjects effect was for goal-type: in general, students use more situations for GC-goals than for GN-goals, and no interaction between goal type and the pre-tests emerged. This lack of effect is due to the small amount of variation in the strategy used to solve GN-goals. Most students solve them by constructing two situations that differ in terms of the feature in question.

Graphical consequence goals: multiple & single case

When different types of GC-goal are distinguished, different approaches to the goals are highlighted by the pre-tests. Table 5 indicates the between-subjects and within-subjects effects.

Table 5: Comparing GCM and GCS goals: between-subjects effects and within-subjects effects.

Pre-test(s)	Between-subjects effects
PFT	$F(1, 25) = 4.43$, $p<0.05$
EFT by PFT	$F(1, 25) = 9.38$, $p<0.01$
EFT by GREA	$F(1, 25) = 4.83$, $p<0.05$
PFT by GREV	$F(1, 25) = 4.47$, $p<0.05$
EFT by PFT by GREV	$F(1, 25) = 6.94$, $p<0.02$
EFT by GREA by GREV	$F(1, 25) = 6.55$, $p<0.02$

Pre-test(s)	Within-subjects effects
goal-type	$F(1, 25) = 42.98$, $p<0.001$
EFT by PFT by goal-type	$F(1, 25) = 8.55$, $p<0.01$
PFT by GREA by goal-type	$F(1, 25) = 4.55$, $p<0.05$
EFT by PFT by GREV by goal-type	$F(1, 25) = 7.15$, $p<0.02$
EFT by GREA by GREV by goal-type	$F(1, 25) = 8.52$, $p<0.01$

Again, those with high scores on the EFT and PFT use the fewest number of situations for GC-goals. For within-subjects effects, a different pattern emerges to that found for S- and G-goals. There is an interaction of PFT by GREA by goal-type, and this indicates different approaches to the two types of goal.

For GCM goals, those that score low on the GREA but high on the PFT use the fewest situations (see Table 6). For GCS goals, these are the students that use the most situations. GCS goals can be solved by using one situation that indicates the sentential information being applied to the graphical situation. Alternatively, they can be solved by constructing multiple situations that explore the constraints. Those students that effectively exploit graphical abstractions in solving the GCM problems seem to maintain this strategy when a shorter solution is available. The GREA low scoring subjects are those that are poorer at solving problems where one model can be constructed. These are exactly the GCS-goals. The current analysis suggests that these students have learned a coping strategy for such problems which is not efficient, but it does at least achieve the result, and being flexible in using graphical representations enables the development of this strategy. Students who are GREA low but do not learn

Table 6: Mean number of situations for goal type by PFT and GREA.

GCM-goal	PFT Lo	PFT Hi
GREA Lo	13.40	9.86
GREA Hi	12.33	10.89
GCS-goal	PFT Lo	PFT Hi
GREA Lo	1.15	2.00
GREA Hi	1.71	1.21

this coping strategy—due to being less able to exploit graphical abstraction—do not experience this interference effect on GCS problems. Those students that score highest on both scales seem to choose optimal strategies: their proofs for the GCM goals are not so short, but when they come across GCS goals, they can recognise and use a more appropriate strategy. This reflects flexible use of the graphical abstraction facilities and recognition of differing proof constraints.

Conclusions

The current study replicates the transfer effects of learning logic in a multimodal environment to other reasoning domains, but the strategic differences previously observed do not emerge in the same way. This is due, in part, to differences in the teaching method which provides extra encouragement for using sentential representations to solve logical problems. This gives rise to three types of strategy: sentential, graphical abstract and graphical concrete, which form a continuum in terms of the extent to which they utilise HP's graphical abstraction facilities.

Distinguishing different goal types in HP offers a window into flexible strategic change and the inter-relationship of multiple psychometric measures. No *one* psychometric captures strategic variation in using graphical abstraction in HP, but combinations reflect the options open to students. If psychometrics index cognitive style, then cognitive style dictates the development of strategies in solving multimodal problems. It is this chaining that reflects the observations of strategic variation and change being related to purported measures of 'spatial ability'. A student's propensity for solving problems in a certain way is tempered or licenced by their ability to use representations to achieve the goal. Some students develop strategies that counteract their difficulty with particular problem types, but these can then result in an inflexibility of approach. Those students who seem flexible in their strategies in HP do not necessarily use these strategies optimally, but their ability to switch strategy according to the problem's constraints makes up for this.

The highlighting of strategic variation and change in the current study enables a recharacterisation of the psychometric measures which takes into account the computational features of the task. Thus, HP offers a window into what the psychometrics mean from the computational perspective. Because HP is based on a principled (if highly abstract) theory of reasoning, its categories can be applied to understanding performance on tests such as the GRE. For example, 'splitting into cases' is something that has to be achieved in reasoning, whether in a formal domain, or in more informal problem solving.

Acknowledgements

This research was supported by ESRC research studentship R00429634206, by ESRC's Centre Grant to HCRC, and by a grant from the McDonnell Foundation's CSEP Initiative. The third author is an EPSRC Advanced Fellow.

References

Barwise, J. & Etchemendy, J. (1994). *Hyperproof*. Stanford: CSLI Publications.

Bennet, C., Haglund, B., Westerståhl, D. & Sönströd, C. (1997). *En introduction till första ordningens språk.* University of Gothenberg.

Cronbach, L. J. & Snow, R.E. (1977). Aptitudes and Instructional Methods: A Handbook for Research on Interactions. Irvington: N.Y.

Cox, R., Stenning, K., & Oberlander, J. (1994). Graphical effects in learning logic: reasoning, representation and individual differences. *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (237-242). Georgia: Lawrence Erlbaum Associates.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Guilford, J.P. (1980). Cognitive styles: what are they? *Educational and Psychological Measurement*, 40, 715-735.

Jonassen, D. H. & B. L. Grabowski (1993). *Handbook of Individual Differences, Learning, and Instruction.* Hillsdale: Lawrence Erlbaum Associates.

Kyllonen, P. C. & Lohman, D. F. (1983). Individual differences in solution strategy on spatial tasks. In R. F. Dillon & R. R. Schmeck (Eds.) *Individual Differences in Cognition Volume 1*. New York: Academic Press.

Mates, B. (1965). *Elementary Logic.* Oxford: Oxford University Press.

Monaghan, P. (1998). Holist and serialist strategies in complex reasoning tasks: cognitive style and strategy change. Research Paper EUCCS/RP-73. Edinburgh: University of Edinburgh, Centre for Cognitive Science.

Monaghan, P. & Stenning, K. (1998). Effects of representational modality and thinking style on learning to solve reasoning problems. *Proceedings of the 20th Annual Conference of the Cognitive Science Society of America* (716-721). Madison: Lawrence Erlbaum Associates.

Oberlander, J., Cox, R., Monaghan, P., Stenning, K., & Tobin, R. (1996). Individual differences in proof structures following multimodal logic teaching. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, (201-206). La Jolla, CA: Lawrence Erlbaum Associates.

Roberts, M. J. (1998). Individual differences in reasoning strategies: a problem to solve or an opportunity to seize? In G. d'Ydewalle, W. Schaeken, A. Vandierendonck & G. De Vooght (Eds.), *Deductive reasoning and strategies*. Mahwah: Lawrence Erlbaum Associates.

Salthouse, T. A., Babcock, R. L., Mitchell, D. R. D., Palmon, R. & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence*, 14, 187-230.

Stenning, K., Cox, R. & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, 10, 333-354.

Witkin, H., Oltman, P., Raskin, E. & Karp, S. (1971). *A manual for the embedded figures test.* Palo Alto: Consulting Psychologists Press.